

A Report from the American Academy of Microbiology

MICROBES

MICROBES

THE GLOBAL GENOME QUESTION:

Microbes
as the Key to
Understanding
Evolution
and Ecology

MICROBES

MICROBES



AMERICAN
SOCIETY FOR
MICROBIOLOGY

Copyright © 2004
American Academy of Microbiology

1752 N Street, NW
Washington, DC 20036
<http://www.asm.org>

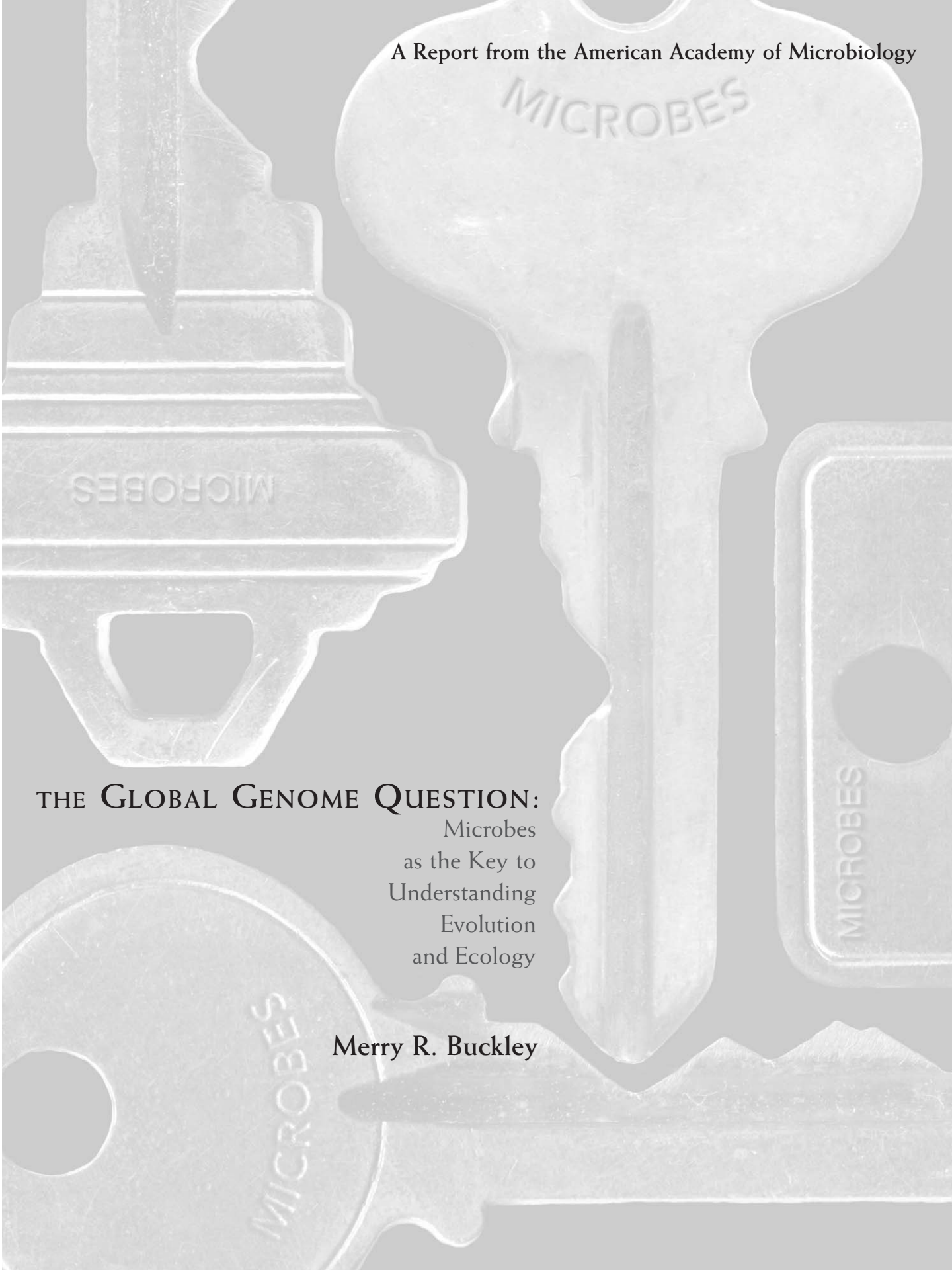
This report is based on a colloquium, "The Global Genome Question: Microbes as the Key to Understanding Evolution and Ecology," sponsored by the American Academy of Microbiology and held October 11-13, 2002, in Longboat Key, Florida.

The American Academy of Microbiology is the honorific leadership group of the American Society for Microbiology. The mission of the American Academy of Microbiology is to recognize scientific excellence and foster knowledge and understanding in the microbiological sciences.

The American Academy of Microbiology is grateful for the generous support of the following organizations:

National Science Foundation
U.S. Department of Energy
NIDCR, National Institutes of Health
The Ellison Medical Foundation

The opinions expressed in this report are those solely of the colloquium participants and may not necessarily reflect the official positions of our sponsors or the American Society for Microbiology.



A Report from the American Academy of Microbiology

THE GLOBAL GENOME QUESTION:

Microbes
as the Key to
Understanding
Evolution
and Ecology

Merry R. Buckley

**Board of Governors,
American Academy of Microbiology**

Eugene W. Nester, Ph.D. (Chair)
University of Washington

Kenneth I. Berns, M.D., Ph.D.
University of Florida

James E. Dahlberg, Ph.D.
University of Wisconsin, Madison

Arnold L. Demain, Ph.D.
Drew University

E. Peter Greenberg, Ph.D.
University of Iowa

J. Michael Miller, Ph.D.
Centers for Disease Control and Prevention

Stephen A. Morse, Ph.D.
Centers for Disease Control and Prevention

Harriet L. Robinson, Ph.D.
Emory University

Abraham L. Sonenshein, Ph.D.
Tufts University Medical Center

David A. Stahl, Ph.D.
University of Washington

Judy A. Wall, Ph.D.
University of Missouri

Colloquium Steering Committee

Edward F. Delong, Ph.D. (Co- Chair)
Monterey Bay Aquarium Research Institute

David A. Relman, M.D. (Co-Chair)
Stanford University School of Medicine

W. Ford Doolittle, Ph.D.
Dalhousie University, Halifax, Nova Scotia, Canada

Karen E. Nelson, Ph.D.
The Institute for Genomic Research

Richard J. Roberts, Ph.D.
New England Biolabs

Carol A. Colgan
American Academy of Microbiology

Colloquium Participants

Siv G.E. Andersson, Ph.D.
Uppsala University, Sweden

Martin Blaser, M.D.
New York University School of Medicine

Brendan Bohannon, Ph.D.
Stanford University

Sallie (Penny) Chisholm, Ph.D.
Massachusetts Institute of Technology

Frederick M. Cohan, Ph.D.
Wesleyan University

Thomas P. Curtis, Ph.D.
University of Newcastle, United Kingdom

Edward F. DeLong, Ph.D.
Monterey Bay Aquarium Research Institute

W. Ford Doolittle, Ph.D.
Dalhousie University, Halifax, Nova Scotia, Canada

Katrina J. Edwards, Ph.D.
Woods Hole Oceanographic Institute

T. Martin Embley, Ph.D.
Natural History Museum, London, United Kingdom

Theresa Gaasterland, Ph.D.
The Rockefeller University

John F. Heidelberg, Ph.D.
The Institute for Genomic Research

Matthew D. Kane, Ph.D.
National Science Foundation

Eugene Koonin, Ph.D.
National Library of Medicine,
National Institutes of Health

William F. Martin, Ph.D.
Heinrich-Heine University, Dusseldorf, Germany

DeEtta K. Mills, Ph.D.
Florida International University

Karen E. Nelson, Ph.D.
The Institute for Genomic Research

Howard Ochman, Ph.D.
University of Arizona

David A. Relman, M.D.
Stanford University School of Medicine

Richard J. Roberts, Ph.D.
New England Biolabs

Gabrielle Rocap, Ph.D.
University of Washington School of Oceanography

Francisco Rodriguez-Valera, Ph.D.
University Miguel Hernandez,
San Juan de Alicante, Spain

Thomas M. Schmidt, Ph.D.
Michigan State University

Melvin I. Simon, Ph.D.
California Institute of Technology

David M. Ward, Ph.D.
Montana State University

Ronald Weiner, Ph.D.
University of Maryland

Stephen Zinder, Ph.D.
Cornell University

Executive Summary

A colloquium was convened in Longboat Key, Florida, in October 2002, by the American Academy of Microbiology to discuss the role of genomic techniques in microbiology research. Research professionals from both academia and industry met to discuss the current state of knowledge in microbial genomics. Unanswered questions that should drive future studies, technical challenges for applying genomics in microbial systems, and infrastructure and educational needs were discussed. Particular attention was focused on the great potential of genomic approaches to advance our understanding of microbial communities and ecosystems. Recommendations for activities that might promote and accelerate microbial genome science were identified and discussed.

Microbiology has always advanced in tandem with new technologies. Beginning with the first observations of microscopic organisms with early microscopes in the 17th century, the tools and methods for studying microbes have continually evolved. Slowly at first, and now with startling speed, scientists have developed increasingly complex and informative tools for analyzing the functions, interactions, and diversity of microorganisms. Today, genomic technologies are revolutionizing microbiology. Genomics employs all or part of the genome to answer questions about an organism and represents a generic tool that can be used to dissect any or all living cells. In this report, the term “genomics” includes structural genomic methods that focus on the determination of genomic sequence and higher order structural features, as well as functional genomic methods, which focus on the activities and products encoded by the genome.

To date, microbial genomics has largely been applied to individual, isolated microbial strains, with the results extrapolated to the wider world of microbial diversity. We are now presented with an opportune moment to move beyond studies of single isolates and to apply genome sciences directly to the study of microbial communities. It is now possible to adapt genomic tools and approaches to more realistic models of genome evolution and ecology involving natural microbial communities. Microbial communities are formed by organized groups of microbial species, each having different, often complementary functions or activities. In aggregate, the microbial community has emergent properties greater than the sum of its individual members. Outside the laboratory, virtually all microorganisms exist in complex assemblages, in which they exchange genetic material, nutrients, and biochemical signals with one another. While analysis of individual strains has been a highly profitable enterprise, greater strides can

now be made by focusing attention on microbial communities. These are the entities that encompass the bulk of microbial interactions, evolutionary processes, and biogeochemical activities, with resulting immense impacts on human health and the entire planetary biosphere.

The natural microbial world can be viewed as a landscape of genes and genome ecology, in which organisms exchange genetic information and co-evolve with one another, shaping themselves and the biosphere over time. Microbial genomic evolution is crafted in microbial communities through the dynamic interplay of mutation, genetic drift, gene transfer, and natural selection. As it is currently envisioned, the application of genomic approaches to the study of microbial communities, i.e. “community genomics” or “meta-genomics,” entails large-scale sequencing of pooled, community genomic material, with either random or targeted approaches, assembly of sequences into unique genomes or genome clusters, determination of variation in community gene and genome content or expression over space and time, and inference of global community activities, function, differentiation, and evolution from community genomic data.

With the aid of genomic techniques, scientists are poised to answer fundamental questions about the nature of microbial communities and the processes that shape and sustain them. Although there are few limits as to the phenomena that can be explored using these tools, certain areas of research deserve particular attention, due to their fundamental importance for understanding microbial life and due to their relative accessibility, given the current state-of-the-art. Outstanding questions about diversity and its generation and maintenance, ecosystem and community stability, and the relative significance of gene transfer in microbial communities need to be addressed.

Genomic techniques are a powerful set of methods, but there are certain technical hurdles to overcome before these techniques can be universally applied. Among these hurdles are the challenges of coordinating productive research programs centered around significant and tractable biological questions and applying appropriate and cost-effective technologies to answer them. One of the biggest challenges is the difficulty of coping with the tremendous complexity of microbial communities and their habitats and the difficulties in measuring all relevant biotic and environmental variables. Certain technical problems, such as identifying minority populations, deciphering diverse chromosome structures, and de-convoluting complex genome assembly problems,

all need to be tackled to accelerate progress in microbial community genomics. In light of these technical challenges, establishing the proper biological and environmental contexts for genomic studies and developing new technology platforms and strategies are high priorities.

Microbial genome science can best be advanced by adopting multiple strategies and by addressing multiple levels of complexity in study systems. Investigations that focus on more simplistic model microbial communities will facilitate methods development. Less complex model systems will also allow multiple research groups to coordinate their work. Several recommendations are made with respect to the optimal attributes of model systems, and a few examples of such systems are explored. At the same time, genomic technologies are mature enough now to decipher the genomic characteristics of more complex microbial communities. Genomic investigations of microbial communities of global ecological, medical, or industrial importance should also begin immediately. Since genomic approaches generate vast amounts of data, improvements in information technology, database architecture, and data management strategies will significantly accelerate scientific progress. Outstanding unanswered questions include the amount and nature of within-population genetic diversity, the dynamics of genome evolutionary processes, and the levels of gene transfer in microbial communities.

A number of well recognized educational needs exist with respect to microbial genome science and are particularly critical for community genomics. These include cross-disciplinary training at the graduate and post-graduate levels, in fields that meld biology and computer science or mathematics.

Microbial genomics holds great promise for improving our world. By enabling a predictive understanding of the effects of perturbation on the microbial communities that impact human health and the environment, genomics could hold the key to treating diseases and managing the precious natural resources and processes that sustain life on this planet.

Introduction

It is astonishing but true—microbes are responsible for maintaining life on Earth. By driving global cycles of oxygen, carbon, and other essential elements, microbes have created the atmosphere, soil, and sediment that support the remarkable diversity of animals and plants that have existed for millions of years. Microbes have played an essential role in the creation of the fertile land that sustains crops and livestock and, therefore, sustains human populations. Even within our own bodies, bacteria are necessary for digestion and absorption of nutrients and for educating our immune systems. Life on earth has thrived because of the activities of microorganisms, and in their absence that life would cease to exist.

Despite the pivotal roles they play in life on Earth, microbes are only beginning to be well understood. From the beginning, the study of microbes has been dependent on advances in technology, and discoveries in microbiology come only at the pace allowed by the methods available. Microbiology was born less than 350 years ago when Antonie van Leeuwenhoek ground glass into magnifying lenses and identified “animalcules” thriving in almost every environment he could imagine. Then, in the 1860s, Louis Pasteur showed that these organisms can cause disease and can be purged from food and medical instruments using heat. Ferdinand Cohn discussed the role that microorganisms play in cycling natural elements in 1872. Robert Koch grew bacteria on solid media 10 years later, allowing bacteria and fungi to be cultivated and studied more easily in the lab. In 1924, Albert Jan Kluyver wrote about the common metabolisms of many different microbes and was the first to point out that life on earth would be impossible without their actions.

In this past century, the development of molecular biology allowed great advances in understanding microorganisms. In 1944, microbiologists Oswald Avery, Colin MacLeod, and Maclyn McCarty showed that deoxyribonucleic acid, DNA, held the information that drives the activities of living things and is the basis of heredity. The polymerase chain reaction (PCR), first conceived in 1983, enabled scientists to take a small amount of genetic material from microbial cells and amplify it for faster, more accurate analysis. Since then, molecular biology has evolved rapidly and scientists have developed a range of powerful techniques for uncovering the activities of microscopic organisms, allowing microbiologists to pinpoint the causes and identify the cures of elusive diseases, to better exploit the beneficial microbes used in food and industrial processes, and to characterize the myriad organisms that cycle global nutrients.

In recent years, scientists have opened a new window on the world of microbes: genomics. The genome of an organism is all of the DNA or genetic material that the organism holds. It is the set of instructions that directs the growth, metabolism, and reproduction of every living creature. In short, the genome is a blueprint for life. “Genomics” rose to prominence with the first description of a complete genome sequence of an independent life-form in 1995. The term describes the study of all or part of the genome to answer a question about an organism. Genomic techniques are powerful, and thus far they have enabled scientists to explore the metabolic capabilities and genetic identities of many microbial strains, leading to insights about the role those organisms play in disease and environmental processes. In this report, the term “genomic techniques” includes approaches that entail genome sequencing and analysis as well as functional approaches, in which the production of gene products in individuals or microbial communities is monitored.

Molecular methods for studying microbes, including genomic techniques, have allowed researchers to examine many genes and functions, but our ability to understand the microbial world has been limited by the difficulties inherent in studying microbes in their natural setting: microbial communities. Microbial communities are formed by organized groups of microbial species each having different, often complementary, functions or activities. In aggregate, the microbial community has emergent properties greater than the sum of its individual members. Outside pure laboratory cultures and exceptional symbioses, virtually all microbes exist in communities. Whether in soil, water, attached to environmental surfaces, or in the gut of a human being, microorganisms live in close proximity to other microbes, and the implications of this closeness are profound. For example, microbes in communities exchange nutrients, taking the substances they require and excreting those for which they have no use. These metabolic activities alter the microenvironments in which microbes live, so microbial interactions can define the impact of a community on its environment. The spatial arrangement of cells in a community can amplify microbial interactions and may be responsible for many of the important processes communities carry out. Community-level processes may also be influenced by cell-to-cell signaling or quorum sensing—enabling many microbial communities to take a census of their own numbers and growth state and to respond to their environment in a coordinated fashion. In dense microbial communities, genetic material may be exchanged, allowing the various members to acquire and relinquish different metabolic capabilities and growth characteris-

tics. Microbial community members can be highly diverse, including members from each of the three domains of life and can display metabolic capabilities of every sort. Alternatively, they may consist of only a few types of members living in close association and relying on simple associations for their needs and sustenance.

The result of dynamic community interactions at genomic, metabolic, and structural levels is an interacting network of microbial species whose collective functions are often greater than the sum of its parts. For example, certain microbial communities are capable of breaking down hazardous chemicals that individual species cannot degrade. Likewise, consortia of bacteria acting together are required to satisfy the nutritional requirements of many larger organisms. Microbial communities are potent, and they accomplish many of the functions that are necessary for higher organisms to exist and for sustaining our planet's habitability.

Microbial communities are many and varied, and their interactions, diversity, and structure make these assemblages complicated systems indeed. This complexity makes it difficult to tease out the function of individual microbes in their natural settings when communities are examined as a whole. Given these difficulties, why should scientists study microbes in communities? The answer is obvious: microbes are inextricably linked to, and defined by, the communities in which they are found. To extract individual members from a community, thrust them into controlled, aseptic laboratory medium, and monitor their behavior in completely foreign nutrients and physical circumstances is of dubious utility to understanding microbes in the real world. Microbes do not exist alone; they exist in communities. Understanding the role that microbes play in the world requires that we understand them at the community level. Hence, efforts to describe the activities of microorganisms should focus on studying them within the context of communities wherever possible.

The pressures of community living, including population dynamics, recombination, and lateral gene transfer, are emerging as critical drivers of ecosystem evolution. By using genomic techniques to approach these phenomena, microbiologists can now build on the vast data resources from genome sequencing projects and carry the science into a new era. In this new stage, genomics will be used to describe the evolutionary forces that drive the microbial communities that impact human health and the environment. Thus, among the challenges before us is the need to determine how the techniques, tools, and perspectives of genomics can be most effectively used to understand better whole microbial communities—as single units of study.

The Genomic Landscape on Earth Today: Its Evolution, Diversity, and Dynamics

It is instructive to view the biological world through the lens of the genome. Most enduring ecological changes are mediated at the level of genes or groups of genes, where the forces of natural selection are brought to bear. Hence, the biosphere can be thought of as a genomic landscape, where the interactions between genomes and the environment dictate the outcomes of evolution and the formation and stability of the biosphere. This view of the world has been referred to as “gene ecology.”

Genomes evolve over time, changing in subtle or radical ways, constantly adapting to the surrounding environment. When two organisms with identical genomes are separated and exposed to different selective pressures over many generations, their genomes can accumulate genetic differences, causing the organisms to diverge and resulting in a branch-like pattern of evolution. The divergent evolution of previously identical genomes is one way in which diversity can be created. Alternatively, new genomic combinations can evolve through lateral gene transfer, by which a cell can acquire genes from its neighbors or lose genes rapidly. Hence, genomes can evolve gradually through vertical transmission of mutations, gene duplications, deletions, and rearrangements. Alternatively, they can evolve more suddenly and sporadically via horizontal transfer of genetic information between different microbial species.

Understanding the different mechanisms of genome evolution offers interesting insights into the process. For example, it is possible to view evolution from the perspective of the organism that gains and loses genes and functions or from the perspective of the gene or group of genes, which can evolve gradually over time or increase their influence by moving to new hosts.

The structure of the genome—meaning the organization of genes with respect to one another—may also play an important role in the evolution of microbial systems. It has been found that the genomes of closely related organisms can actually exhibit a great deal of variability in their structure. That variability can increase over time. Genome heterogeneity has been documented in strains of *Thermotoga*, for example, and in a number of pathogens, such as *Helicobacter pylori*, but it is unknown what precise role this heterogeneity plays in the evolution of communities of organisms.

Lateral gene transfer is the exchange of genes between microbial strains, and intragenic recombination refers to the re-arranging of elements within a gene. Together, these processes, known as gene transfer, have been

found to be significant driving forces in the evolution of microbial systems. This contradicts prior beliefs, which held that these processes were important only in a limited number of microbial habitats. Gene transfer allows individual organisms access to novel genetic material, and hence, to novel metabolic and structural capabilities that may offer an evolutionary advantage. Lifestyle and local habitat affect the rates and mechanisms of gene transfer, and some bacterial species appear to absorb genes from other organisms so readily that they are likened to a “gene vacuum.” For example, *Helicobacter pylori* readily incorporates genes from the DNA pool to which it is exposed. It can take up genes from a plethora of different organisms, acquiring the capability to metabolize previously excluded substrates or tolerate adverse conditions. It is worth noting, however, that named prokaryotic species have been shown through population genetics analysis to include many discrete subpopulations, each of which might have the properties of species. Until we better understand how genomic variation is organized into populations, we cannot be certain that our impressions of genomic variation within a species are correct. In the case of related

Applying Genomics to Microbial Communities—The Most Important Unanswered Questions

Advantages of Taking a Community-Based Approach

Community-based genomics approaches offer a key to understanding the world of microbes. Previously, meaningful research in genomics often demanded that individual microbial community members be cultivated in the laboratory, in isolation from the forces and relationships of the communities that sustained them in their natural habitats. The requirement of cultivation has imposed a number of constraints on research, not the least of which is the problem of culturing microbes that have yet to be cultured. Currently, we are only able to study thoroughly at best one percent (1%) of known bacteria and archaea, as the culture requirements are either unknown or impractical to implement for the vast majority of prokaryotic microbes. Other limitations include the difficulty of extrapolating the results of cultivation-based investigations to the wide world outside

Genomes evolve over time, changing in subtle or radical ways, constantly adapting to the surrounding environment.

groups of organisms that freely exchange genes, the concept of a group-wide genome is useful. The presence of agents that mediate gene exchange, namely phages and plasmids, also has an influence on the rates and mechanisms of gene exchange. Until now, the barriers to studying gene transfer and to quantifying the effects of gene transfer on evolution and community dynamics have been daunting. Genomics-based community approaches are well suited to exploring these gaps in the current understanding of the microbial world, and the results of such investigations promise to be fruitful.

the lab and differences between the pure-culture behavior of microbes and the behavior of microbes in communities. Community-level genomics approaches have enabled scientists to take a step back from the approaches that involved dissecting communities into their individual member cells and enabled them to step simultaneously into new realms of research where the overarching processes, interactions, and relationships at work in communities can be more directly observed. It is also important to note that the most significant outstanding questions in microbiology are related to ecological and evolutionary processes, which can only be studied in the relevant context of the communities and environments in which they occur.

Genomics and other molecular methods are extremely powerful, and they eliminate many of the problems associated with cultivation-dependent techniques. An important and critical place still exists, however, for the continued exploration of microbial cultivation, which plays a role in understanding the metabolic and structural feats of which microbes are capable. Genomics could even help to guide these investigations by allowing researchers to identify the organisms appropriate for cultivation and the likely combination of culture techniques that would enable their growth in the lab.

Box 1: Community Genomics and the Metagenome

“Community genomics” refers to the application of genomics tools and techniques to the study of entire microbial communities as a single unit. This approach begins with pooled DNA or RNA representative of the whole community and involves the creation of an inventory of genes of the community in aggregate, using random shotgun sequencing methods, or targeted sequencing of large genome fragments. It may involve the assembly of individual genome sequences that correspond to single strains, species, or genomic variants. In a manner analogous to efforts in single organism genomics, community genomics projects seek to define the expression patterns and function of the gene pool under a variety of conditions. Variation in content or expression in relation to time, space, and perturbation help reveal the biology and evolution of the community as a whole. In discussing community genomics, the term “metagenome” has been coined to refer to the entire set (or inventory) of genes that belong to a given community.

Genomics approaches applied to communities offer unique promise for answering many outstanding questions—from identifying the minimum required gene set in defined communities to fundamental questions about the evolution of all ecological systems. General areas of inquiry where genomics can be of particular value in leading to new discoveries include:

- Cataloguing diversity of microbial communities;
- Understanding how genomic diversity relates to species and/or functional diversity;
- Deciphering how genome composition and dynamics relate community stability;
- Elucidation of the nature extent, and relative significance of lateral gene transfer; and
- Mapping the functions, dynamics, and interactions of individuals and communities within specific environmental settings.

Outstanding Questions, Unique Opportunities

Diversity

It has long been acknowledged that the functional diversity of the microbial world is staggering. From the termite gut to deep-sea hydrothermal vents to unseen groundwater plumes of noxious chemicals, microbes carry out forms of metabolism and respiration that are entirely beyond the capabilities of larger organisms. Within the last 20 years, scientists have found that the genetic diversity of microbes is also vast. Microbes occupy most of the branches of the tree of life, representing the majority of the genetic potential in the biosphere. However, this extensive functional and genetic diversity, even within defined microbial communities, has confounded both efforts to characterize the forces that drive diversity and to identify and interpret diversity patterns. Scientists using genomic techniques have begun to tease apart the rules of microbial diversity, but many questions remain.

Molecular analyses of microbial communities based on conserved genes and associated phenotypic properties suggest that diversity is structured in discrete populations that organize along environmental gradients. The term “ecotype” has been suggested in reference to these ecologically specialized populations. For example, closely related “ecotypes” of the most abundant photosynthetic organism on Earth, marine *Prochlorococcus*, are partitioned along light and nutrient depth-related gradients in the sea. The levels of molecular resolution are currently too low for most microbial groups to investigate whether other organisms (and their genomes) are organized into species-like ecotype populations. Genomic approaches can address the issue of whether individual microorganisms form populations of organisms of like function or genetic relatedness. If so, efforts can also be made to go beyond statistical characterization of these groupings to determine the underlying biological processes that might drive the phenomenon. Evolutionary theories, such as periodic selection theory, can be used to formulate testable hypotheses to begin to understand these processes. On the other hand, ecological theories developed for plants and animals might not apply to microbial populations, especially if lateral gene flow is frequent, so that theory-independent approaches to discovering the processes influencing how microbial genomes evolve must also be considered.

Efforts can also be directed toward understanding the degree of variability in microbial community diversity and the changes in the relative abundance of genes and genomes that accompany community perturbation.

Often taken for granted, the correlation between genomic diversity and functional diversity within communities has not been thoroughly explored. With the powerful genomic tools currently available, the time is ripe to substantiate such intuitive connections with experimental results, and model and predict genome evolutionary dynamics with robust theory.

The boundaries of microbial diversity are also largely unknown and could be effectively resolved using genomic techniques. For example, the minimum number of genes necessary for survival as a self-sustaining organism, outside of a symbiotic relationship, is unknown and could be investigated using these tools. Moreover, it is unknown whether the minimum gene set differs based on the metabolic lifestyle of a given microbe. For example, do organisms that utilize organic carbon have a smaller or larger minimum gene set than organisms that fix inorganic carbon? On a community level, it is unknown whether communities operate close to the minimum gene set or whether they retain many more genes than the minimum, and what the biological mechanisms for retaining those extraneous genes might be. Neither is it known whether a community's minimum gene set varies across time and space.

Many ecologists agree that species diversity is correlated with ecosystem stability, but a consensus has certainly not been reached. By using genomic techniques to examine the effects of system perturbations on genome content in microbial communities, microbiology can inform these and other debates relevant to classical ecology.

There are multiple plausible theories about the nature of microbial diversity. Under the ecotype concept of diversity, variation within each ecologically distinct population (an ecotype) is purged recurrently by natural selection. Such ecotypes would have the quintessential properties of species, in that each is genetically cohesive and each is irreversibly separate from all others. These ecotypes would be expected to coexist over evolutionary time, such that each would eventually be discernible as a separate sequence cluster. Alternatively, in a species-less concept of diversity, what limits the variation within an ecologically distinct population is not recurrent selection but the population's longevity. Under the species-less concept, new ecologically distinct populations would arise at a high rate, owing to horizontal transfer, but populations would also become extinct at a high rate. These two models make very different predictions for discovering ecological diversity through molecular methods. Under the ecotype concept, there is a one-to-one correspondence between ecologically distinct populations and sequence clusters; under the species-less concept, the correspondence is many-to-

one. Genomic approaches can help to determine which of these models most appropriately describes diversity within microbial communities.

Stability and Adaptation to Perturbation

Community stability is another topic where genomics tools can be put to use. In combination with phylogenetic and genomic approaches, researchers can begin to dissect the mechanisms of community adaptation to perturbation by separating the effects of (1) shifts in population structure, (2) changes in gene expression patterns, and (3) gene transfer. Since genomic techniques can examine individual microbial groups at work together, the effects of perturbation and stress on communities can be gauged and extrapolated to the communities that impact the biosphere. This will allow researchers to understand whether different stressors induce dormancy, community composition shifts, ecotype extinctions, stepwise or continuous changes, or other phenomena.

The question of whether similar niches support similar microbial occupants could be addressed using genomics as well. Of particular interest is the level of resolution at which these occupants share genetic or functional similarities. For example, in a given type of niche that is found in many different locations, the occupants may belong to the same "species," genus, or higher phylogenetic cluster.

Genomics can help to bring microbiology into a new, predictive stage. Genomic tools could potentially be employed to construct a framework by which certain features of an organism's genome might be predicted from the sequence of its small subunit rRNA gene (16S or 18S) and the biochemical and physical characteristics of the environment from which it was drawn. Alternatively, predictions can be made about the effects of perturbation on microbial communities based on the findings of genomic surveys. This is an especially critical capability in light of its potential impact on managing certain human diseases, which can be seen as disturbances of the delicate balance that humans usually maintain with their microbiota or the microbiota in their environment. (Examples of diseases that may be provoked or propagated by microbial community disturbances include inflammatory bowel disease, periodontal disease, "bacterial vaginosis," and the overgrowth syndromes associated with antimicrobial use.) Perturbations of microbial communities that come into contact with the human body can lead to disease, and if predictions could be made about the effects of these perturbations on human tissues, then therapies could be more effectively targeted to counteract or prevent the deleterious effects.

Lateral Gene Transfer

Lateral gene transfer has recently been found to be more prevalent in microbial communities than previously thought. Hence, it is reasoned that gene transfer must play some significant role in community dynamics and evolution, but it has been difficult to explore this role with the techniques available to date. Genomics allows experimental approaches and hypothesis testing that was impossible with older methods, and concerns relevant to lateral gene transfer are particularly amenable to testing by these methods.

Although it is widely acknowledged that lateral gene transfer is a large determinant of microbial community function and evolution, the agents responsible for gene transfer (plasmids, phages, transposons, integrons, and other elements) have received relatively little attention. Not only are there numerous gaps in our knowledge about the agents of gene transfer, but the factors that govern this phenomenon are also little understood. The rules and trends underlying gene transfer are ripe for exploration using genomic techniques. With the introduction of genomic techniques, systematic community-based surveys of mobile and transmissible genetic material are possible. As exchangeable packets of information, these elements can be seen as “sub-genomic organisms” or “subcellular genetic entities.” Simple physical separation techniques can isolate phages, plasmids, and other genetic fragments and sequencing these elements would be very revealing and of only moderate cost.

The range of rates of gene transfer that can be observed in different types of organisms or in different environments is unknown, as is the relative importance of the various mechanisms responsible for gene transfer. Limits on mechanisms of gene transfer are poorly understood. For example, it is not known whether exchange is limited more by phylogeny or proximity. Other unknowns include the average quantity of genetic material that is exchanged by these transfer mechanisms, and whether there are “species cores” of genes that are excluded from lateral exchange and recombination.

Gene transfer is not only a way of acquiring capabilities, but it may also be a way of “cleaning house,” ridding cells of unnecessary genetic burdens. Whether or not there are common strategies in balancing the ebb and flow of lateral gene transfer to optimize genome structure is unknown. Similarly, the ability or inability to give and receive genetic material in this way may be a determinant of evolutionary change, either enabling cells to take advantage of changing environmental conditions or driving the extinction of cells that cannot adapt.

Each of these lines of inquiry can be effectively investigated using modern genomic techniques. Genomic research investigating lateral gene transfer will require scientists to generate new theories about life in microbial communities.

Questions about gene expression can be addressed using genomic techniques, as well. Although gene expression has been explored in a myriad of cultivation-dependent studies, genomic investigations into the nature of gene expression in a microbe’s “natural” context, in the presence of its fellow community-members, promises to deliver a new understanding of how and when microbes put their inherited genetic codes into action. In sequencing the genomes of isolated microbes, it is not uncommon to find that a third of the genes have no known function. Through genomics investigations that examine the expression of genes in a community environment, including coordinate gene expression, great progress can be made in identifying the roles of these previously unknown, unidentified genes. Other outstanding questions that are ripe for exploration using genomic techniques include the power of environmental feedback on microbial community composition, function, and the tempo and mode of genome evolution.

Questions Relevant To All Biology

The implications of applying genomic techniques to microbial communities are not limited to the realm of microbiology; rather, they resonate in every aspect of biological science. Genomics can be applied to microbial ecosystems to test hypotheses of population genetics for larger organisms, for example, and could have implications for understanding the dynamics of genes in all ecosystems. Outstanding questions with respect to evolution, including whether diversity is organized by periodic selection, can also be approached by use of these techniques and can inform discussions of evolution with respect to larger organisms. Finally, the results of genomic studies that examine the influences of microbial community members on other community members via cell-to-cell interactions can be extrapolated to understand better the interactions between the cells of multicellular organisms.

Issues and Challenges in Applying Genomics to Communities

Applying Ecological Principles to Microbial Communities

In the visible world, living things are found in arrangements called ecosystems, in which the various members establish a niche of specialty, interact with other members of the ecosystem, and arrange themselves in patterns in which the various species utilize the resources necessary for survival. Scientists have studied ecosystems for many years, uncovering the rules that govern them and the emergent properties that come from combining many diverse organisms into a single, functioning whole. Microbial communities help structure ecosystems on both small and large scales. Hence, experimentation in microbial community ecology may be used to inform the realm of classical ecology, and genomic techniques may be a particularly important set of tools in making these links. For example, the phenomenon of invasive species in large ecosystems has a parallel in microbial communities: the

dialog between classical ecology and microbial ecology will often lead to a better understanding of our world, and attempts should be made to translate the results in one field in order to illuminate the other. Interactions like these could lead to new insights into the dynamics of both the biological systems that we can see and those that work on the microscopic level.

Technical Issues, Challenges

Genomics as applied to microbial communities is a relatively new field. A number of challenges need to be faced and resolved in order to advance most effectively knowledge in the arena of microbial genome science.

COMPLEX POPULATIONS

Although genomics has introduced new ways to explore the complexity of microbial communities, that complexity may yet confound some lines of inquiry. Microbial communities exhibit a wide spectrum of intricacy. Some communities, intracellular parasites or single-species symbiont communities, for example, present relatively simple systems for study, and the range of the interactions at work is comparatively small.

Scientists have studied ecosystems for many years, uncovering the rules that govern them and the emergent properties that come from combining many diverse organisms into a single, functioning whole.

introduction of a new gene into the community. Until now, the available techniques in microbiology have allowed researchers to track only the invasion of a given microbe into a community, not the intrusion of the unique genetic material that that organism introduces to the system. Genomic techniques now allow for tracking of genes, and the results of investigations into the movement of an invasive gene through a species, population, or community can contribute to our knowledge about the movement patterns of invasive plants, insects, or animals through the environment.

In other areas, the parallels between classical ecological and microbial community principles break down, and rules that apply to one apparently do not apply to the other. For example, the concept of biogeography, the governance of ecosystems according to climactic or geographical environs, may or may not have a parallel in the realm of microbial communities. Microbial distributions may instead be ruled predominantly by biogeochemistry, and the availability of niches and nutrients appropriate for community survival. In the areas where microbial ecology and classical ecology do not meet, genomic information may not be useful in informing the principles of ecology. However, a continued

However, other microbial communities are comprised of webs of inextricable interactions, in which nutrients, chemical signals, genes, and behavioral interactions are exchanged between different populations in an almost hopelessly complex network. Genomic work on these complex communities can be very difficult indeed, since complexity confounds reproducibility. Genomic surveys should be designed to allow for a range of expected complexities in the community being studied, without biasing the experimentation based on the level of diversity anticipated by the researcher. The best research will operate at several levels simultaneously, including the genomic level (in which surveys and some complete sequencing will be carried out), the gene expression level, the metabolic level, and the physical-chemical level, and the population level.

LEVEL OF RESOLUTION

The interpretation of genomic data, and specifically the question of the level of detection achievable with these methods, remains a challenge in the field. Microbial communities may be comprised of thousands of distinct microbial taxonomic groups roughly equivalent to "species," with most of these contributing to the overall function and character of the community. However, the

vast majority of these groups are also be present at a very low frequency, contributing small signals to the genomic data. In other experiments where genomics is used to compare communities, before and after treatment, for example, or from different areas of an experimental site, small community differences may be critical to the results of the experiment. Without careful planning and experimental design, small, critical groups and small, critical differences between communities may go undetected. This could hinder the interpretation of experimental results, leading to false conclusions and improperly directing future research efforts. Some prior knowledge of the diversity of the system at hand is required to tailor the questions that drive such genomics experiments, allowing proper experimental design to identify minor populations and subtle differences.

CHROMOSOMAL COMPLEXITY

In most of the bacteria currently in culture, the majority of the genome is kept in the form of a single, usually circular chromosome that bears the genes necessary for survival. It is well known that the information found in extrachromosomal elements (such as plasmids), i.e., the genetic information that is not included in this chromosome, also contributes to the biology many microbes. It is becoming increasingly clear, however, that many bacteria, such as *Vibrio cholerae*, bear more than one chromosome. Moreover, eukaryotic microbes, such as fungi and protozoa, like their larger cousins the plants and animals, bear many chromosomes, each of which carries a distinct set of genes, and mitochondrial DNA, which carries yet more genetic information. The fact that the genome of a given microbe cannot always be found on a single, intact circle of DNA can be a source of difficulty in deconvoluting genomic data from complex communities. In the future, researchers must design experiments appropriately to overcome the problems associated with accessing information on multiple chromosomes.

GENOME ASSEMBLY IN COMPLEX COMMUNITIES

The amount of sequence data that is accumulated from even a modest genomics investigation can be enormous, and the computational handling requirements related to analyzing microbial community data can be difficult to overcome. When reconstructing the genomes of organisms from community DNA samples, confounding factors such as insertions, deletions, polymorphisms, and replicons that do not contain a phylogenetic marker like small subunit rRNA genes, can cause inappropriate alignments, creating genomic sequences that are actually agglomerations, or chimeras, of genome fragments from very different organisms. Simulation studies subjecting known whole genome sequences to shotgun cloning and reconstruction indicate that most mixtures of genomes from different microbial species can be assembled en

masse. However, the difficulties related to within-population insertions, deletions, polymorphisms, are potentially substantial. The levels and types of genetic heterogeneity, and relative richness and evenness of different genotypes, will have major impact on attempts at whole community genome assembly. More data need to be collected on microbial genomic diversity within and between populations, to understand, model and solve these problems.

Normalizing genomic data to remove the dominant genomes in the sample may be a key to accurate genome assembly. Normalization may cause the investigator to lose the very rare organisms in the sample (those present at a frequency of approximately 1 in 10,000 or less), but this approach can allow the assembly of the average, or mean, community. Certainly, the inability to detect minor community members is not limited to genomics techniques; it is a problem in many applications in microbial community analysis. It should be remembered that genomics does not overcome all of the limitations of earlier methods and is not the definitive set of tools in understanding community function.

LATERAL GENE TRANSFER

Questions remain about the rate and extent of lateral gene transfer in microbial communities. As a result, it is unknown how stable genomes really are. The composition of genomes may respond dynamically to evolutionary pressures over relatively short time periods, complicating the interpretation of complete genome sequences. A genome sequence may be a "snapshot" of the genes that a given organism carries and it may be subject to relatively rapid change due to lateral gene transfer in response to environmental conditions and other selective pressures. For example, it was recently found that, although the genomes of two strains of *Escherichia coli*, K12 and the deadly O157:H7, share a similar structure and are of similar size, a total of 1,915 genes are carried by one strain and not the other. Hence, as much as a third of the genome of *E. coli* apparently is exchangeable and dynamic. Further work to delineate the significance of lateral gene transfer to microbial genomes should help to illuminate this issue, guiding researchers in ways to conduct meaningful genomic surveys that take the influence of gene transfer into account.

Mobile Elements in Ocean Microbes

Research has shown that as much as 10% of the DNA in the world's oceans is found in phages, indicating that the potential exists for extensive lateral gene transfer among planktonic microbes. By quantifying, sequencing, or otherwise characterizing phage genes in seawater filtrate or any other habitat, researchers could take a "snapshot" of the genes that are on the move at a given point in time. These results could then be compared to the genes that are moving at different times or in different locations. This kind of investigation is already being directed at the phage community within human the human gut.

COMPLEX ENVIRONMENTS

Many environments where microbial communities are found are not amenable to experimentation with direct genomic techniques in which DNA from the entire microbial community is extracted and used in subsequent analyses. Complex environments that contain soluble organic materials and inhibitory compounds are particularly difficult. Soil, for example, is a notoriously troublesome environment in which to apply genomics or other molecular techniques. Composed of a heterogeneous mixture of organic and mineral, soluble and insoluble materials, soil contains a number of substances that may interfere with the extraction process. Furthermore, soluble materials may co-elute with the extracted DNA, contaminating the sample and inhibiting PCR, cloning reactions like ligation, or other processes in the analysis. Although advances in extraction and purification techniques are moving the field forward every year, difficulties with environmental matrices persist and must be managed in many genomics investigations of complex environments.

GENOMES IN CONTEXT

Microbial community genome dynamics can only be understood in the natural biotic and environmental setting in which it evolves. Genome diversity and structure, in part, captures natural historical events. Approaches that integrate genomic techniques with other levels of information include systems modeling, which could facilitate a predictive understanding of communities and the functions of individuals. Mass spectrometry and nuclear magnetic resonance imaging could also prove useful, as could geochemical techniques that can quantify physical and chemical parameters and track them over time. Ecological and environmental data and modeling are also natural complements to microbial community genomic data.

Challenges in Community Genomics

Although genomic techniques can be a very powerful set of tools in investigating the function, dynamics, and structure of microbial communities, there are tradeoffs in using any given analytical approach. The drawbacks to using genomics, including problems of reproducibility and expense, should be carefully considered prior to selecting the methods to be used in addressing a particular question.

Ensuring reproducibility remains a challenge in applying genomic techniques to microbial communities, as microbial communities are extremely complex, and it has long been acknowledged that complexity confounds reproducibility. It may be unknown, for example, how representative a sample from a given community is. In the case of an unrepresentative sample, one that bears little resemblance to the larger community, subsequent analyses can produce results that have little relevance to the community as a whole. Issues of scale are also of major concern when sampling communities that interact across distances of only a few micrometers or less. Genomic surveys of microbial communities should include sufficient independent replicate samples and analyses to ensure that the results of the experiment can be reproduced faithfully. The question remains, however, how much replication is sufficient? This can only be answered on a case-by-case basis, and researchers need to closely scrutinize replication in their experimental design.

Another significant problem with current genomic techniques is their relative insensitivity in detecting minor community members. This can pose a serious constraint on experimental results, and should be considered carefully. Since minority community members can sometimes have disproportionately large functional impact, this is an important area to address.

Educational Requirements and Information Dissemination

Genomic techniques have enabled great advances over the past ten years, granting us insights into the functions carried out by microbial communities and how these communities impact our world. In the future, progress will continue and new discoveries will reveal hidden truths about microbial communities, allowing medicine to derive novel health strategies, enabling ecologists to predict the effects of perturbation on the processes that control greenhouse gases, and permitting engineers to optimize the degradation processes that remove toxins from contaminated water and soil. In order to move the field forward quickly, however, a number of needs must be fulfilled. Scientific needs

exist, including better tools for data management and methods for analyzing rare members. In addition, education and training efforts need to be directed toward preparing scientists for optimal use of genomic techniques.

Development and Characterization of Model Communities and Ecosystems

Model microbial communities and ecosystems can help focus the efforts of separate research groups and facilitate a more thorough understanding of complex assemblages. By providing fine-scale details about representative ecosystems, models allow researchers to construct informed, testable hypotheses about their favorite non-model communities, thus guiding research into productive areas. Moreover, model systems allow researchers to focus resources and expertise on a few useful systems rather than on many different communities that may or may not inform an understanding about other systems.

It is widely agreed that model microbial communities in natural contexts should be studied in order to move our understanding of microbiology forward. Microbiologists need to form a consensus, not only on the value of particular model systems, but also on the criteria for selection of more complex communities. Applying genomics techniques to suitable model ecosystems could prove highly profitable and provide, for example, detailed information on the diversity, stability, and occurrence of lateral gene transfer in a few systems that can then inform an understanding of other microbial systems.

Criteria for Model Selection

A primary criterion for a model community or ecosystem is that it be of fundamental scientific value, theoretical interest, or practical importance. These models can span the spectrum from the simple to the complex. Some simple systems may be amenable to laboratory study and shared between different laboratories. Other model ecosystems might, instead, gather the focused interest of different investigators around a specific geographic site or habitat. Models for use in understanding microbial communities should not only prove instructive in the realm of microbiology, but they should have relevance to other scientific or social issues. In other words, a model should have “real-world” relevance. The more groups, or stakeholders, that have an interest in gaining knowledge about the system in question, the more suitable it is as a model.

Models should also be robust, allowing researchers to test a wide range of hypotheses.

An ideal model microbial community would be derived or would exist in an environment that is amenable to experimentation. Environments can hinder research by virtue of their complex matrix (i.e., soils), lack of ready access (i.e., remote deep sea environments), extreme conditions (i.e., volcanic sulfur springs), or any combination of these challenges. Efforts should be made to ensure that the environment from which a model system is derived does not prevent thorough experimentation or long-term study.

The “physiology” of a model system should be relatively simple and/or informative, in order to facilitate modeling efforts and systems analysis approaches. By studying communities that carry out fundamental physiological processes, research can be focused on those phenomena that are most pervasive in microbial communities and have the biggest impact on human health and the environment.

Criteria for Model Selection

- Fundamental Value or Practical importance
- Tractable
- Robust
- Amenable environment
- Simple physiology
- Stable

Finally, a model community or ecosystem should be stable in terms of structure and function, in order to ensure that the results of research are repeatable over time and in different laboratories.

Scale of Model Systems

A balance must be struck between the complementary advantages of using simple communities with relatively few members and using large, complicated ones with many members. Although a model microbial community should be tractable and have a small enough number of members so that the system can be approached using genomic techniques and other methods, this requirement does not mean that model communities should be simple in all cases. The question of the appropriate scale of complexity for a model community can be answered by determining what hypotheses are being tested and/or what critical problem needs to be solved. These are the ultimate drivers that will determine what value a particular model has for advancing knowledge and human welfare. A range of

- 16 Importantly, a model microbial community should be tractable to continued study. The complexity or number of different kinds of organisms should not be so great that genomics tools, in their current configurations, cannot be used to evaluate the dynamics of the system.

complexities should be studied that span a spectrum from simple two member consortia to complex multi-species communities. Only by approaching a variety of levels of complexity can the field be advanced rapidly and limitations tested and overcome.

Simple communities, those with few members, offer some advantages to the researcher in these early stages of development. The tractability and reproducibility of these systems may provide early understanding the dynamics of microbial communities. A community of limited size may also simplify genomic analyses, enabling researchers to reconstruct the genomes of the community with greater ease.

More complex communities, on the other hand, are probably more representative of the way life is for most microbes on this planet. They are also more likely to be self-organizing and self-sustaining, qualities that make them a good model of ecosystems and of the earth's biosphere. For this reason, developing approaches for deciphering the dynamics of complex communities represents a high priority endeavor. It should be the ultimate goal, and an immediately proximal activity, in microbial community genome analysis.

Examples of Model Communities

Model systems are needed in order to make faster progress in understanding disease, the environment, and our global ecosystem. As a starting point for discussions about which models should be recommended to the microbiology community, a few examples of suitable models were discussed.

MICROBIAL MATS

Thick layers of microbes, called microbial mats, often grow where geochemical conditions are appropriate, usually in thermal, hypersaline, or brackish marine environments. These communities have been shown to be stable, allowing repeatable experimentation. The chemical environments within microbial mats are amenable to study by microensors, presenting definable, predictable gradients, and the number of organisms usually found in these communities is well within the range that is reasonable for careful genomic analyses.

SYMBIOTIC ASSOCIATIONS

Symbiotic systems, in which organisms apparently cooperate to the benefit of all members, are also suitable as model systems. Symbioses are frequently very simple communities of as few as one or two microbial groups, and the physical boundaries of these communities are usually well defined. However, the range of interactions in these communities can be limited, having little bearing on larger, more complex systems. For example, symbiotic microbial

communities may not be subject to invasion by non-native microbes, which limits the lessons these systems can offer for communities that are vulnerable to invasion.

OPEN OCEAN GYRES

Open ocean gyres, vast expanses of the oceans that represent one of the biggest ecosystems on the planet, may present a suitable model system for addressing several outstanding questions about microbial community structure and function. Dominated by small planktonic bacteria and archaea, microbial communities in open ocean gyres have been found to have levels of microbial diversity that make them eminently tractable systems. Since these communities largely control carbon and energy cycling in the sea, they are of tremendous practical and ecological significance as well.

HUMAN GASTROINTESTINAL TRACT AND ORAL CAVITY

At birth, the human gastrointestinal tract and oral cavity are sterile environments. Upon exposure to the world, these environments are quickly colonized by microorganisms in what may be predictable, sequential succession. As such, the gastrointestinal tract and oral cavity could serve as ideal model systems for studying a number of phenomena, including community establishment and succession. An understanding of perturbation and its effects on community structure in these systems might lead to novel strategies for health maintenance and disease prevention.

MICROBIAL COMMUNITIES RESPONSIBLE FOR CORAL DESTRUCTION

The microbial communities associated with destruction of the world's corals clearly engage a number of concerned stakeholders interested in knowing more about the mode of action, spread, and methods for controlling these destructive assemblages. Undoubtedly, a systematic study of one or more of these communities at many levels of resolution would be a good first step in making these advancements.

Managing Data

Both the power and the puzzle of genomic techniques lie in the vast datasets generated by using these methods. Large amounts of data can be derived even from simple genomic experiments, allowing researchers to reach conclusive, detailed conclusions about their system. Managing, analyzing, and integrating these data, however, remain a challenge for the field. Applying genomics to understanding microbial communities whose structure remains incompletely explored contributes even more complexity to the problem of data management. For example, we do not yet have a proven concept of microbial species, and we do not even know whether microbes fall into groups with

species characteristics. However, if the ecotype concept of microbial diversity proves true, this construct could prove useful for organizing a community's genomic and metabolic diversity. Organisms from different ecotypes would be discernible by the sequence clusters in which they fall, and then differences in genomic content among organisms could be more easily interpreted. For example, genes not shared among different ecotypes would be candidates for being the basis of adaptive differentiation among ecotypes; genes not shared among members of the same ecotype would be ecologically meaningless, caused by random horizontal transfer events. On the other hand, if the species-less concept turns out to be correct, it would be more difficult to determine which gene differences are responsible for ecological diversity within the community. Determining how microbial diversity is organized will facilitate interpretation of genomic data.

A great need exists for better algorithms and computer software to handle the vast amounts of genomic data generated by community genomics applications. Mathematical and informatic formulations for handling sequence and expression data, for example, are scarce to unavailable. Moreover, the expertise for creating these tools is not abundant (See Needs and Education Requirements).

Genomics databases are also lacking in number and quality. Historically, biologists have been ineffective in the "care and feeding" of databases, an unfortunate phenomenon that now extends to poor maintenance and tracking of genomics databases as well. By and large, diminished responsibility for biological databases, including genomics databases, is correlated with a lack of enthusiasm for funding these efforts. The reluctance to support the necessary maintenance of genomic databases must be overcome if progress in genomics is to continue unabated.

Better integration between databases is a worthy goal for informatics work in the future, as new data are generated almost daily. It is particularly important to include relevant physical, chemical, and biological data when constructing these databases. Often, these data are excluded in compiling information about microbial communities, limiting the value of the resulting database for comparing the results of different investigations.

As databases begin to proliferate, and the amount of data collected in investigations of microbial communities continues to grow, making genomics information accessible to the scientific community has become more and more difficult. Efforts need to be focused on informing scientists of the resources available to them and simplifying access to those resources, particularly

online resources, so that information can be effectively disseminated and utilized.

Microbiology has become an interdisciplinary science, and the application of genomics technologies to studying microbes is no exception. Unlike many prior areas of research, however, genomics involves a strong informatics component, requiring microbiologists to take computing and mathematical lessons to heart. Collaborations between biologists, engineers, mathematicians, and professionals in other aligned fields are likely to be the most productive efforts in investigating microbial communities.

Gene Transfer

The fact that so many outstanding questions remain about lateral gene transfer, a phenomenon that could be a significant determinant of community dynamics and evolution, stands in the way of making advancements in understanding life in microbial communities. Clearly, a need exists to answer these fundamental questions. By focusing on questions about the rates, limits, agents, and preferred modes of lateral gene transfer, researchers can build a base of knowledge that can serve as a springboard for new theories and hypotheses about population dynamics, evolution, and adaptation.

Analyses of Individuals

Although the focus of applying genomics in microbiology should be the characterization of microbial communities, a need still exists for identifying and analyzing individual cells. Advancing our ability to accomplish this should not be neglected. A great deal of our current knowledge about the interactions, functional capabilities, and limits of microbial communities has come from studies of individual isolates, studies which have offered views on the finer details of microbial life. Efforts to analyze single cells and microbial community members should move forward in the future, allowing researchers to sequence the genomes of individuals and to observe the activities of those community members *in situ*. The technology to cultivate previously uncultured microbes should also be encouraged, as microbes in cultivation can give concrete evidence of the diversity of metabolisms and functions that microbes are capable of carrying out in their natural environments.

Other Needs

In order to move the field into a predictive stage, wherein the effects of microbial community perturbation on human health and the environment can be anticipated, efforts are needed to conduct coordinated surveillance of the impacts of disturbance on communities at very basic levels.

There is also a need for centralized repositories for archiving raw and processed materials from community analyses, which will allow easier verification and reproducibility of the results from these investigations. A federal facility would seem the best option for archive sponsorship.

The labs involved in microbial genomics are currently operating with almost complete independence, often with little knowledge of the efforts of other research groups—that is, until research results are published in the peer-reviewed literature. In order to maximize the rate of progress in the field, research projects should be conducted in a coordinated fashion, rather than competitively. Microbial genomics laboratories should assemble and participate in a research coordination network, in which the work of individual groups can be cross-referenced with the work of other labs, maximizing the effectiveness of their cumulative endeavors.

Education Requirements

Models, data management, characterization of gene flow, and other issues are pressing topics in optimizing

research in its own right, with its own set of principles, goals, and methodologies.

Cross-disciplinary training is another area where the current educational system may limit the forward momentum of genomics. It is critical to train graduate students and postdoctoral associates in the disciplines that bridge traditional departmental boundaries.

One particularly important need is to train microbiologists, molecular biologists, and biochemists in principles of evolution and ecology. These foundation subjects in general biology are often omitted in these more modern disciplines. A poor understanding of such basic biological laws may seriously reduce our ability to formulate sensible biological questions in microbial ecology and evolution. ASM should ensure that students of microbiology base their views on natural biological laws. Other examples of cross training between disciplines include training in both biology and mathematics, or in molecular biology and computer science. Programs like these, in which students receive in-depth training in aligned fields that have rele-

Cross-disciplinary training is another area where the current educational system may limit the forward momentum of genomics.

the power of genomics, but the education and training of the researchers who will apply genomic methods is paramount. Without sufficient genomics training, effective research would never come about. For example, it has been noted that computing power is probably going to be less of a restriction in dealing with vast amounts of genomics data than the availability of adequately trained scientists. Efforts need to be made by universities, funding agencies, and individual scientists to encourage development of new, interdisciplinary fields, train scientists across traditional departmental boundaries, attract professionals with cross-disciplinary interests, and develop curricula that reflect advances in genomics.

The application of genomics to microbial systems has brought to light novel, cross-disciplinary fields of specialty, including computational biology. Infrastructure, financial support, and educational support are needed to ensure that these areas remain viable disciplines. Moreover, it is important to create the intellectual division between the relatively new fields that have relevance to genomics and their parent disciplines. For example, computational biology is seldom seen as an independent field of research, but is viewed as a subdivision of computer science or biology. Computational biology should be considered a necessary field of

vance in the application of genomic techniques, should be developed and supported at the university, state, and federal levels.

While the educational system acclimates to the current need for scientists trained both in biology and computer science or mathematics, research organizations, including universities, private companies, and other entities, must compete to attract those few individuals who have an interest or are skilled in bioinformatics. A dialog must be established between biology researchers and informaticists to motivate the latter to become involved in genomics research. The mechanisms that encourage these interactions, namely collaboration requirements by funding agencies, are few and far between, and more routes for interactions are needed. Moreover, career promotion mechanisms are needed as incentive for informaticists to stray from the realm of computer science and into the realm of biology. Biology is notorious for failing to adequately recognize the contributions of computer scientists to research programs. This failure needs to be addressed if competent, motivated professionals are to be attracted to the field of genomics.

It is too often the case that undergraduate curricula reflect an out-of-date understanding of biology, and fail

to represent recent advancements in science to students who are interested in the future of research. The role of genomics in the future of science, and of microbiological science in particular, needs to be emphasized, as advancements in our knowledge by using these tools are likely to guide the cutting-edge of research for years to come. For example, biology programs should address the ways in which our understanding and our modes of experimentation in evolution are likely to be affected by genomics. Ongoing partnerships between teachers and laboratories or institutions should be encouraged and redoubled so that the powerful role of genomics in advancing science and broadening our understanding of life can be effectively communicated to students and the public.

Genomics in the 21st Century – What We Can Accomplish

Genomics has revolutionized the way biological science is executed. In microbiology, the tools and techniques of genomics have an unprecedented power to eventually enable the development of a new, predictive science. In the future, advances in genomics will allow scientists to extrapolate microbial community function and the impacts of disturbance on human health and the environment from the results of straightforward analyses. Although a great deal of work stands between us and these future capabilities, the contributions that microbial genomics is poised to provide for medicine and the environment are extraordinary.

Currently, our ability to predict the responses of a single microorganism from the sequence of its genome can best be described as feeble, and our ability to make predictions for an assemblage of multiple organisms is even weaker. However, we can make certain predictions about the inability of communities to carry out certain functions. For example, if genomic methods reveal that an individual or a community lacks *nif* genes (the genes that confer the ability to fix nitrogen), it can safely be predicted that the individual or community is incapable of fixing nitrogen.

In the future, more powerful predictions will be based on the assumption that genomically similar communities have similar dynamics and carry out similar functions. This ecosystem predictive capability will come from detailed work that links an understanding of the genome to an understanding of gene expression, protein function, and complex metabolic networks. In this way, by describing the parts of microbial communities and then extrapolating from the model, scientists can create predictive tools that will allow an under-

standing of ecosystem function. Ecosystem function, which is at the root of human disease as well as environmental stability, has fundamental biological, medical, and environmental relevance.

Predictive tools can enable novel health strategies in medicine. The ability to predict disturbances in the communities present in the human body and in our surrounding environments will allow doctors to avoid or correct the community imbalances that result in disease in a predictable, repeatable fashion that bypasses current trial-and-error approaches. Antimicrobial therapy may be employed in a more discrete, less disruptive, and more informed fashion. In the future, “community diagnostics” will enable better prognostics and therapeutics, all of which will be based upon microbial genomics.

In the environment, advances in genomics will enable scientists to predict the impacts of community disturbances. Of particular interest is the ability to predict the extinction of genes, organisms, and functions that are relevant to environmental health, agricultural production, and global cycles of the elements. Microbial communities could be used as a source of diagnostic signatures for environmental health, enabling better management of our limited natural resources. For example, the microbial communities of rice paddies have an unquestionable impact on world food supplies, and concerns exist about the impacts of community perturbations on the global production of rice. Similarly, the microbial communities of the oceans control global cycles of the elements and strongly influence climate. Complex microbial communities in the environment of the human body are major factors in both health and disease. In the future, using genomic information and analyses, these impacts and influences will be tracked, predicted, and potentially manipulated. A deep genomic understanding of these integrated microbial processes will provide a better understanding of our planet, our interaction with it, and our ability to predict and influence its future behavior.

Recommendations

Enhance research community access to production DNA sequencing facilities

The scientific community needs more open access to production genome sequencing facilities and associated bioinformatics tools. To date, there has been limited access to high throughput sequencing for the general community. Research dollars for specific science projects are often disproportionately redirected to maintain production laboratory infrastructure and overhead. Better mechanisms need to be developed to create more opportunities and less “closed shop” scenarios in the production sequencing laboratories. More open mechanisms for providing production DNA sequencing and open source bioinformatics analyses to the general scientific community need to be available. Models for community sequencing access are already developing, such as the United States Department of Energy’s (DOE) Joint Genome Institute Community Sequencing Program. Even more opportunities like this need to become available in the future, as well as programs that supply the necessary follow-up funding for post-sequence analysis.

Create centers that facilitate research community access to postgenomic analytical capabilities

Centralized infrastructures that can collaborate with the community, and supply large scale “DNA chip” or microarray facilities, proteomic instrumentation and analyses, imaging facilities, advanced fluorescence activated cell sorting facilities, and similar technological platforms are needed. National Laboratories, Centers of Excellence, and National Science Foundation (NSF) Science and Technology Centers represent some of the current mechanisms to create and enhance such structures. Community solutions that facilitate access to genome-based technologies will allow the research community to focus better on merging the necessary disciplines and on the scientific questions.

Identify and nourish strategic collaborations across disciplinary boundaries

The trajectory of the scientific questions posed in the genomic analysis of microbial communities necessitates transcending disciplinary boundaries. Computational biologists and bioinformaticians need to work side by side with environmental scientists. Microbial ecologists need to interface with theoretical evolutionists and population geneticists. Medical microbiologists would benefit from more ecological thinking. The blurring of disciplinary boundaries places well-known stressors on funding, peer review, academic departmental, and tenure granting mechanisms. These challenges need to be faced

openly by academic departments, university administrations, academic journals, and funding agencies. The challenge is to foster integration, while maintaining strong disciplinary foundations and high academic and disciplinary standards and quality. Areas that would stand to particularly enhance the analyses of genomic data in microbial community contexts include, but are not limited to, microbial ecology and population genetics, bioinformatics and environmental science, microbial ecology and medical microbiology, and systems biology and environmental microbiology.

Develop community databases that integrate microbial community genomics with relevant environmental, clinical, geographical, and geochemical data

We lack “biological information systems” that link biological datasets with associated physical, chemical, and other environmental data. Such databases and database standards should be developed. Development of such databases will need to be a large interdisciplinary community effort, requiring the input of computational biologists, population geneticists, environmental scientists, ecologists, and microbiologists. A plan will need to be developed to maintain such inter-operable databases, and this will require time and a broad community input.

Develop computational tools to analyze the complexity at all levels of biological information within and between complex microbial communities

New sorts of computational tools and algorithms will be required to inter-relate the large and complex datasets that will arise from microbial community genome sequence analyses. One particular challenge of significant magnitude will be the assembly of whole genomes from sequences obtained from complex mixtures of highly related organisms. The current DNA assembly strategies and algorithms will need to be modified significantly to meet these new challenges. In addition, new ways to inter-relate genomic, environmental, clinical, biogeochemical, and biogeographical datasets will need to be developed and applied.

Encourage large efforts that focus on specific, naturally occurring model microbial communities and ecosystems

Establishing focused groups of researchers that team to study particular microbial communities or ecosystems has potential to accelerate the development of tools and techniques for genome-enabled microbial community analyses. Specific attributes of model microbial communities might include widespread distribution, geochemical importance, biomedical importance, tractability, or industrial importance. Those microbial communities that have already been

extensively well studied using traditional methods have a built-in infrastructure and historical datasets that would enhance the value of genome enabled approaches. Model communities may come from an environment that is amenable to genomic experimentation and has a relatively simple physiology.

Establish research programs that investigate the process and dynamics of microbial community genome evolution in natural systems

Efforts to establish the relationships among genomic diversity, ecological diversity, and ecosystem dynamics should be encouraged and accelerated. In particular, the processes and products of microbial speciation, diversification and lateral gene transfer events need to be better documented and understood. The rate and extent of interspecies gene flow in microbial communities is largely unknown, and should be investigated in order to understand better the stability of the microbial genome. The relative influence of lateral gene transfer and other modes of genome evolution and diversification need to be much better quantified, modeled, and placed in appropriate ecological contexts.

Develop enabling technologies to dissect more accurately genomic content and dynamics in complex microbial communities.

New and better technologies to dissect the individual and collective genomes found in naturally occurring microbial populations need to be developed. Front end purification techniques including novel cultivation strategies and cell purification strategies such as flow activated cell sorting, coupled with newly developed single genome amplification strategies, could greatly accelerate progress. Downstream analytical strategies for sequencing the collective genomes of whole communities also need development and improvement. In particular, algorithms for de-convoluting intraspecific and within-population genomic polymorphisms, need extensive development and improvement.

Suggested reading

- Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. Delong. 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289:1902–1906.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10988064&dopt=Abstract.
- Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case, and W. F. Doolittle. 2003. Lateral gene transfer and the origins of prokaryotic groups. *Annual Review of Genetics* 37:283-328.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=14616063&dopt=Abstract.
- Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer. 2003. Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology* 185:6220-6223.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=14526037&dopt=Abstract.
- Cohan, F. M. 2002. What are bacterial species? *Annual Review of Microbiology* 56:457-487.
http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12142474&dopt=Abstract.
- Daubin, V., N. A. Moran, and H. Ochman. 2003. Phylogenetics and the cohesion of bacterial genomes. *Science* 301:829-832.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12907801&dopt=Abstract.
- DeLong, E. F. 2002. Microbial population genomics and ecology. *Current Opinions in Microbiology* 5:520-524.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12354561&dopt=Abstract.
- DeLong, E. F. 2002. Towards microbial systems science: integrating microbial perspectives, from genomes to biomes. *Journal of Environmental Microbiology* 4:9-10. [No abstract available.]
- Feil, E. J., and B. G. Spratt. 2001. Recombination and the population structures of bacterial pathogens. *Annual Review of Microbiology* 55:561-590.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=11544367&dopt=Abstract.
- Lerat, E., V. Daubin, and N. A. Moran. 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the *Proteobacteria*. *PLoS Biology* 1:101-109.
http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12975657&dopt=Abstract. Full text at <http://biology.plosjournals.org/plosonline/?request=get-document&doi=10.1371/journal.pbio.0000019>.
- Mann, N. H., A. Cook, A. Millard, S. Bailey, and M. Clokie. 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* 424:741.
http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v424/n6950/abs/424741a_fs.html&dynoptions=doi1074264256.
- Nelson, K. E. 2003. The future of microbial genomics. *Journal of Environmental Microbiology* 5:1223-1225.
[No abstract available.]
- Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299-304.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10830951&dopt=Abstract.

- Rocap, G., F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser, and S. W. Chisholm. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. . 424:1042-1047.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12917642&dopt=Abstract.
- Rodriguez-Valera, F. 2002. Approaches to prokaryotic biodiversity: a population genetics perspective. *Environmental Microbiology* 4:628-633.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12460270&dopt=Abstract.
- Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology* 66:2541-2547.
http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10831436&dopt=Abstract. Full text at <http://aem.asm.org/cgi/content/full/66/6/2541?view=full&pmid=10831436>.
- Sullivan, M. B., J. B. Waterbury, and S. W. Chisholm. 2003. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* 424:1047-1051.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12944965&dopt=Abstract.
- Tamas, I., L. Klasson, B. Canback, A. K. Naslund, A. S. Eriksson, J. J. Wernegreen, J. P. Sandstrom, N. A. Moran, and S. G. Andersson. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 296:2376-2379.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12089438&dopt=Abstract.
- Ward, D. M. 1998. A natural species concept for prokaryotes. *Current Opinion in Microbiology* 1:271-277.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10066488&dopt=Abstract.